

In the format provided by the authors and unedited.

Natural variation in the parameters of innate immune cells is preferentially driven by genetic factors

Etienne Patin^{1,2,3,26*}, Milena Hasan^{4,26}, Jacob Bergstedt^{1,5,6,26}, Vincent Rouilly^{3,4}, Valentina Libri⁴, Alejandra Urrutia^{4,7,8,9}, Cécile Alanio^{1,4,7,8}, Petar Scepanovic^{10,11}, Christian Hammer^{10,11}, Friederike Jönsson^{12,13}, Benoît Beitz⁴, Hélène Quach^{1,2,3}, Yoong Wearn Lim⁹, Julie Hunkapiller¹⁴, Magge Zepeda¹⁵, Cherie Green¹⁶, Barbara Piasecka^{1,2,3,4}, Claire Leloup¹⁴, Lars Rogge^{4,17}, François Huetz^{18,19}, Isabelle Peguillet^{20,21}, Olivier Lantz^{20,21,22,23}, Magnus Fontes^{6,24}, James P. Santo^{4,8,25}, Stéphanie Thomas^{4,7,8}, Jacques Fellay^{1,9,10}, Darragh Duffy^{4,7,8}, Lluís Quintana-Murci^{1,2,3,27}, Matthew L. Albert^{1,4,7,8,9,27*} and The Milieu Intérieur Consortium²⁸

¹Unit of Human Evolutionary Genetics, Department of Genomes & Genetics, Institut Pasteur, Paris, France. ²CNRS UMR 2000, Paris, France.

³Center of Bioinformatics, Biostatistics and Integrative Biology, Institut Pasteur, Paris, France. ⁴Center for Translation Research, Institut Pasteur, Paris, France. ⁵Department of Automatic Control, Lund University, Lund, Sweden. ⁶International Group for Data Analysis, Institut Pasteur, Paris, France.

⁷Laboratory of Dendritic Cell Immunobiology, Department of Immunology, Institut Pasteur, Paris, France. ⁸INSERM U1223, Paris, France. ⁹Department of Cancer Immunology, Genentech, South San Francisco, CA, USA. ¹⁰School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. ¹¹Swiss Institute of Bioinformatics, Lausanne, Switzerland. ¹²Antibodies in Therapy and Pathology, Department of Immunology, Institut Pasteur, Paris, France. ¹³INSERM U760, Paris, France. ¹⁴Department of Human Genetics, Genentech, South San Francisco, CA, USA. ¹⁵Employee Donation Program, Genentech, South San Francisco, CA, USA. ¹⁶Department of Development Sciences, Genentech, South San Francisco, CA, USA. ¹⁷Immunoregulation Unit, Department of Immunology, Institut Pasteur, Paris, France. ¹⁸INSERM U783, Faculté de Médecine, Site Necker-Enfants Malades, Université Paris Descartes, Paris, France. ¹⁹Lymphocyte Population Biology, CNRS URA 1961, Institut Pasteur, Paris, France. ²⁰Center of Clinical Investigations CIC-BT1428 IGR/Curie, Paris, France. ²¹Department of Biopathology, Institut Curie, Paris, France. ²²Equipe Labellisée de la Ligue de Lutte Contre le Cancer, Institut Curie, Paris, France. ²³INSERM/Institut Curie U932, Paris, France. ²⁴Centre for Mathematical Sciences, Lund University, Lund, Sweden. ²⁵Innate Immunity Unit, Institut Pasteur, Paris, France. ²⁶These authors contributed equally: Etienne Patin, Milena Hasan and Jacob Bergstedt. ²⁷These authors jointly directed this work: Lluís Quintana-Murci and Matthew L. Albert. ²⁸A list of members and affiliations appears at the end of the paper. *e-mail: epatin@pasteur.fr; albertm7@gene.com

T cell panel

Phenotypes:

T cells (CD3⁺)

CD4⁺

CD4⁺CD8a⁺

CD8b⁺

DN

In CD4⁺ and CD8b⁺:

Naive CD27⁺CD45RA⁺

CM CD27⁺CD45RA⁻

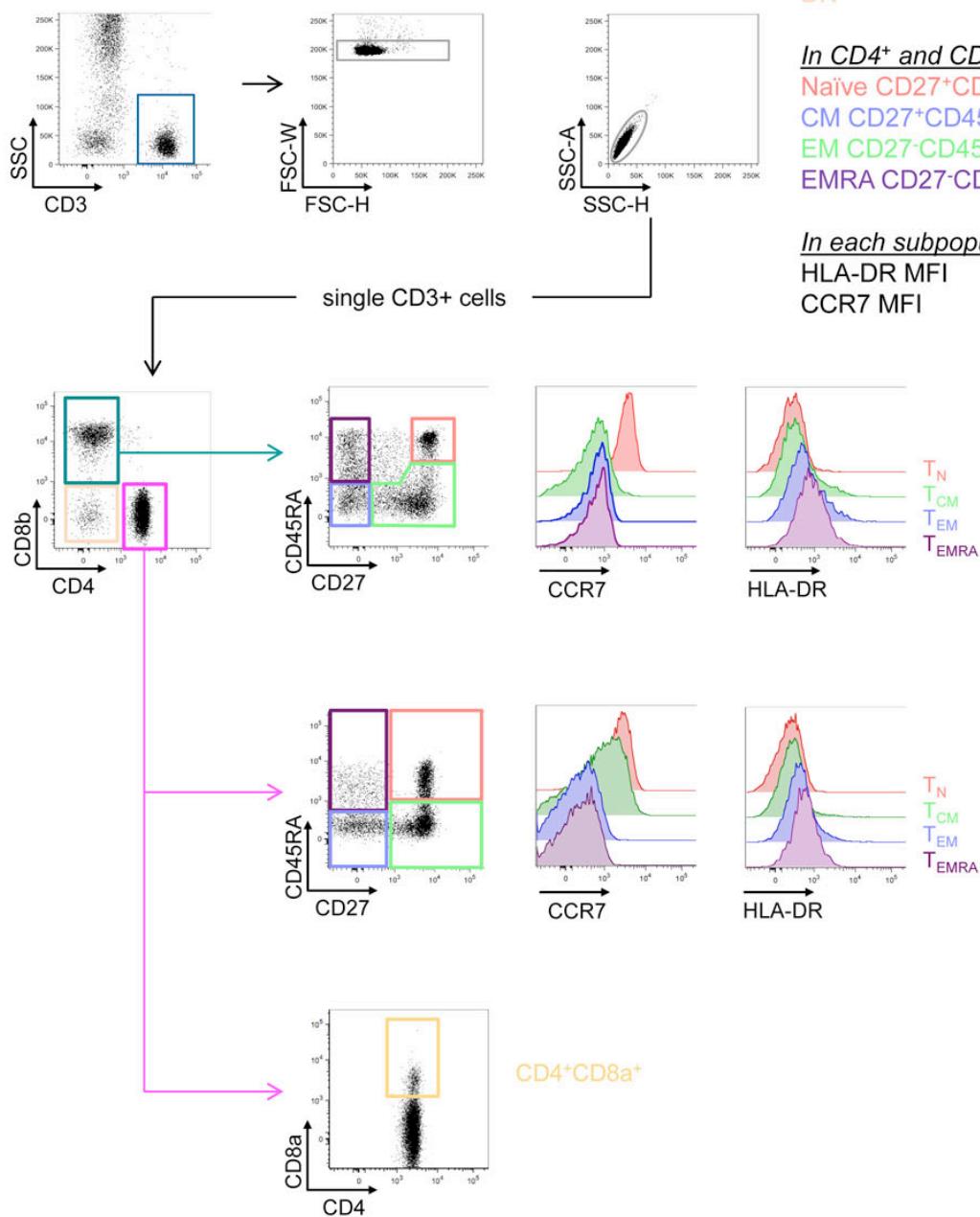
EM CD27-CD45RA⁻

EMRA CD27-CD45RA⁺

In each subpopulation:

HLA-DR MFI

CCR7 MFI



Supplementary Figure 1

Gating strategy for the T cell flow cytometry panel

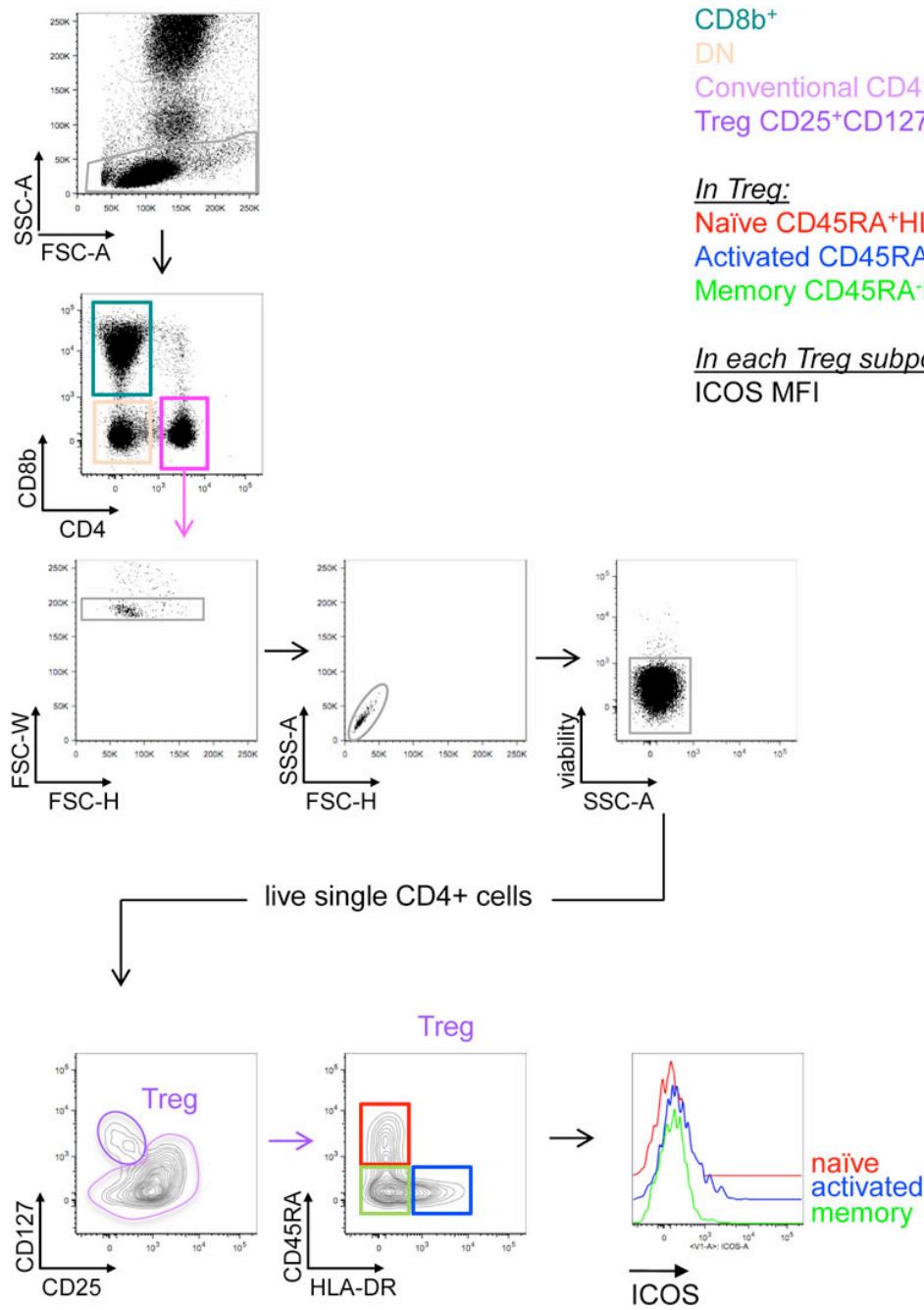
Treg panel

Phenotypes:

CD4⁺
CD8b⁺
DN
Conventional CD4⁺ (CD127⁺)
Treg CD25⁺CD127⁻

In Treg:
Naïve CD45RA⁺HLA-DR⁻
Activated CD45RA⁻HLA-DR⁺
Memory CD45RA⁻HLA-DR⁺

In each Treg subpopulation:
ICOS MFI



Supplementary Figure 2

Gating strategy for the T_{reg} cell flow cytometry panel

MAIT/NKT panel

Phenotypes:

T cells (CD3⁺)

TCR $\gamma\delta^+$

CD8b⁺

CD4⁺

DN

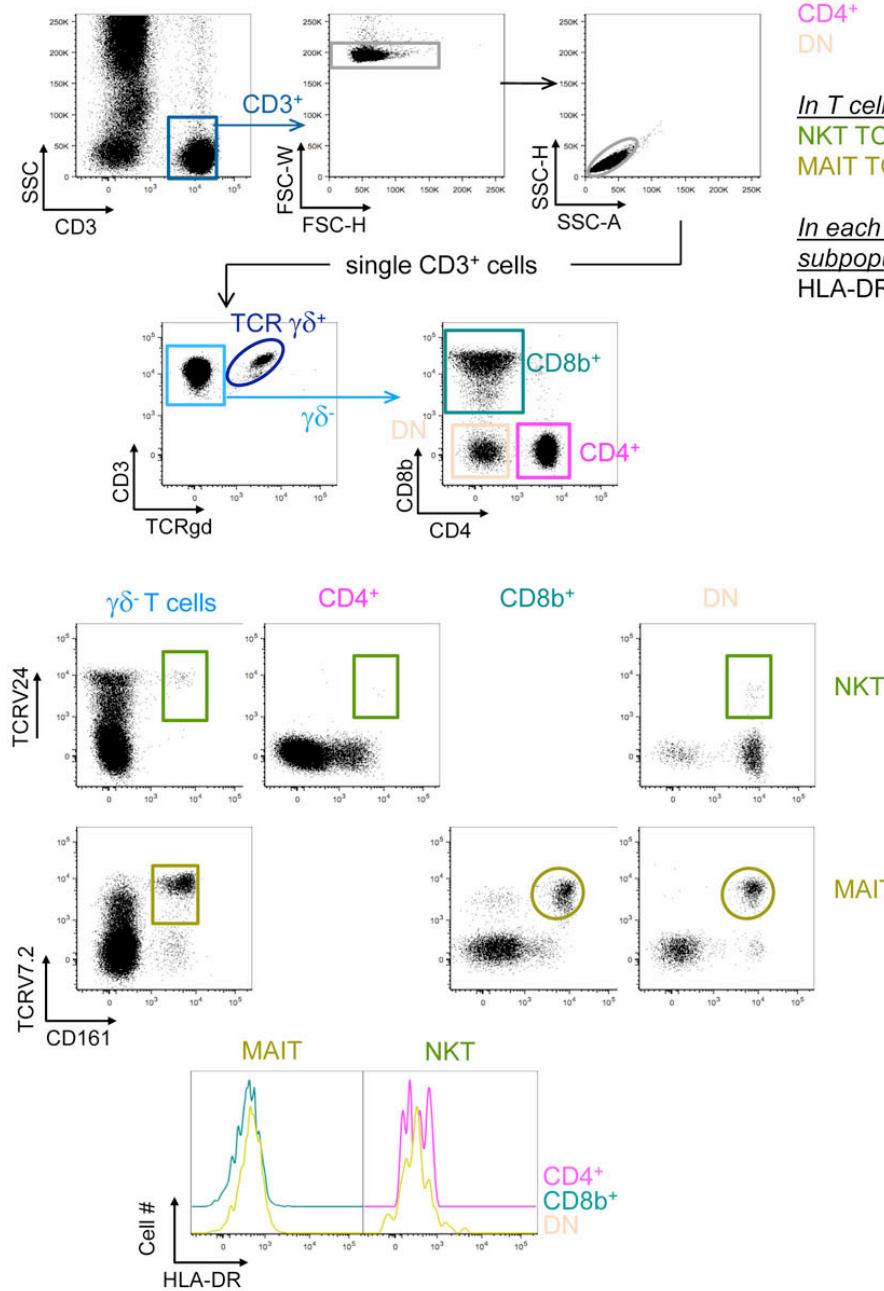
In T cell subsets:

NKT TCRV24⁺CD161^{hi}

MAIT TCRV7.2⁺CD161⁺

In each MAIT & NKT subpopulation:

HLA-DR MFI



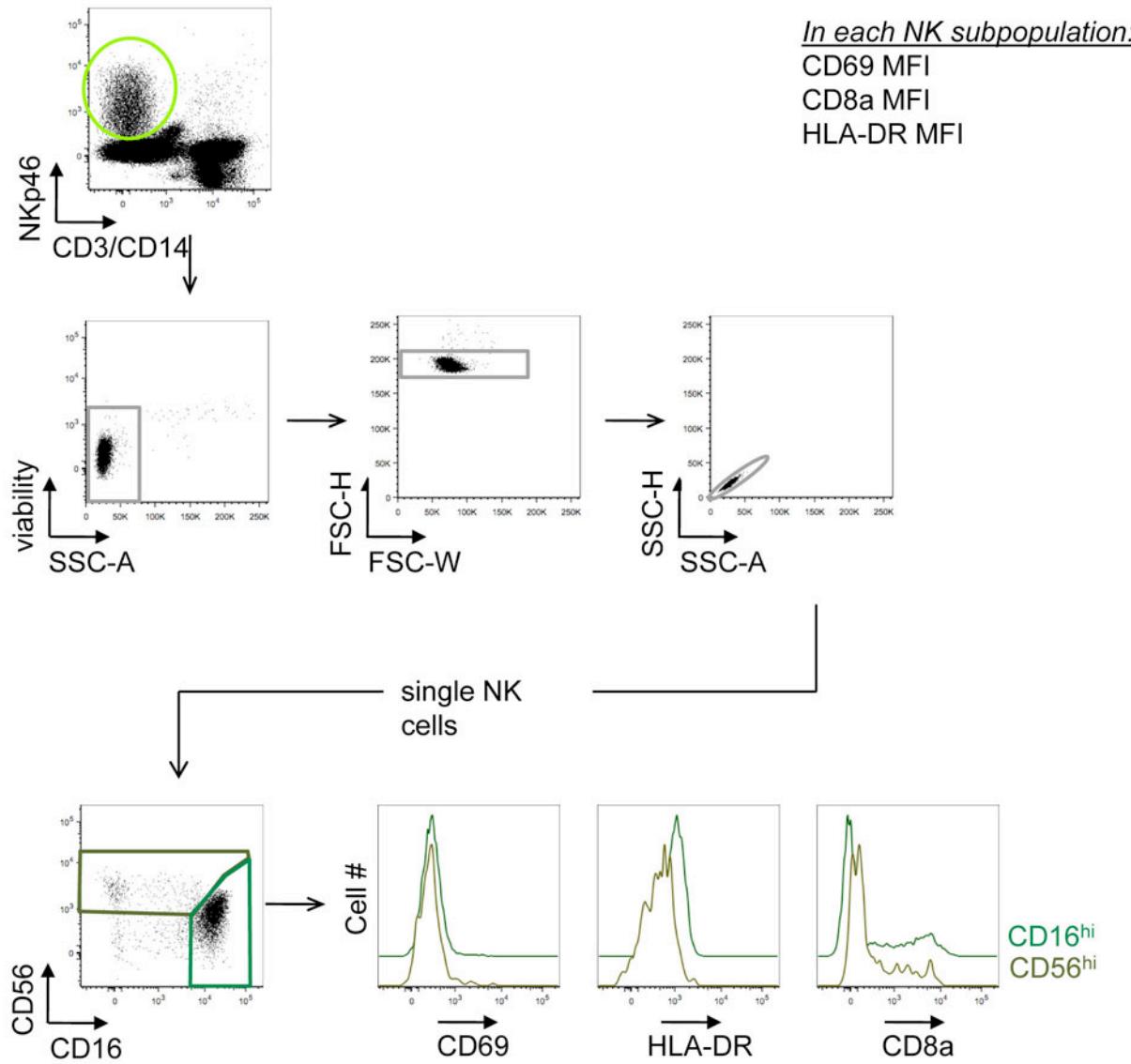
Supplementary Figure 3

Gating strategy for the NKT/MAIT cell flow cytometry panel

NK panel

Phenotypes:

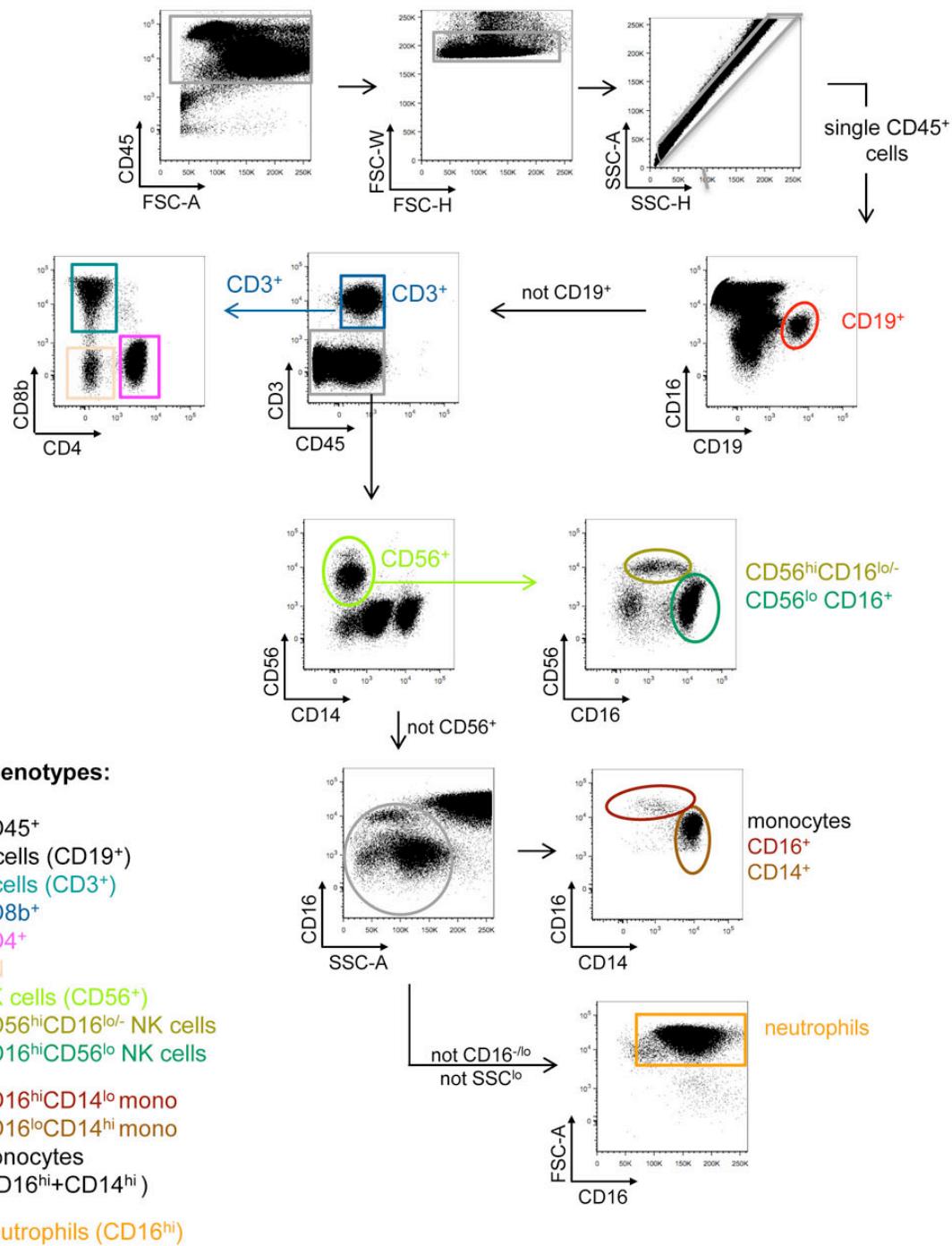
NK cells (NKp46^+)
 $\text{CD56}^{\text{hi}}\text{CD16}^{\text{lo/-}}$
 $\text{CD16}^{\text{hi}}\text{CD56}^{\text{lo}}$



Supplementary Figure 4

Gating strategy for the NK cell flow cytometry panel

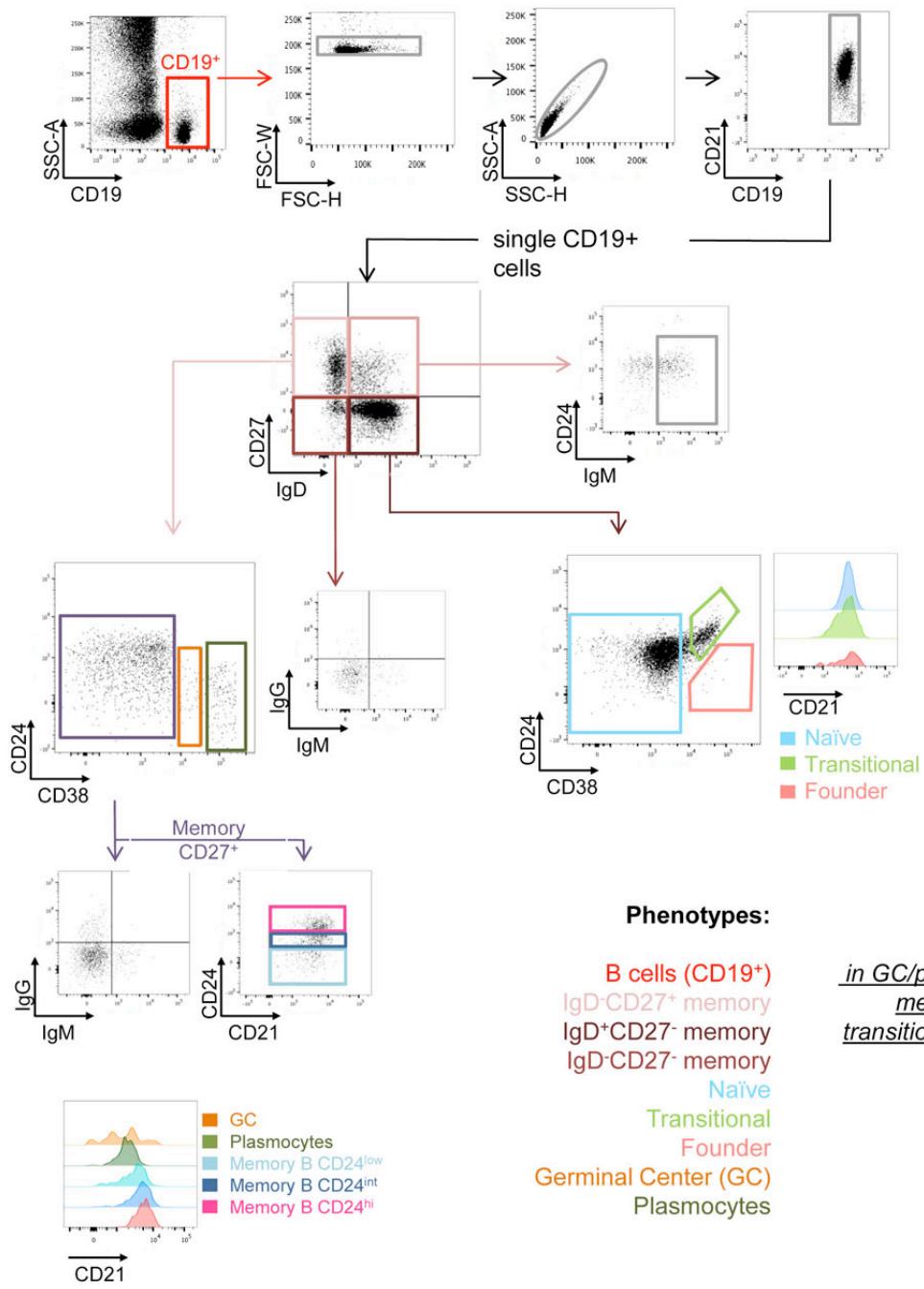
Lineage panel



Supplementary Figure 5

Gating strategy for the lineage cell flow cytometry panel

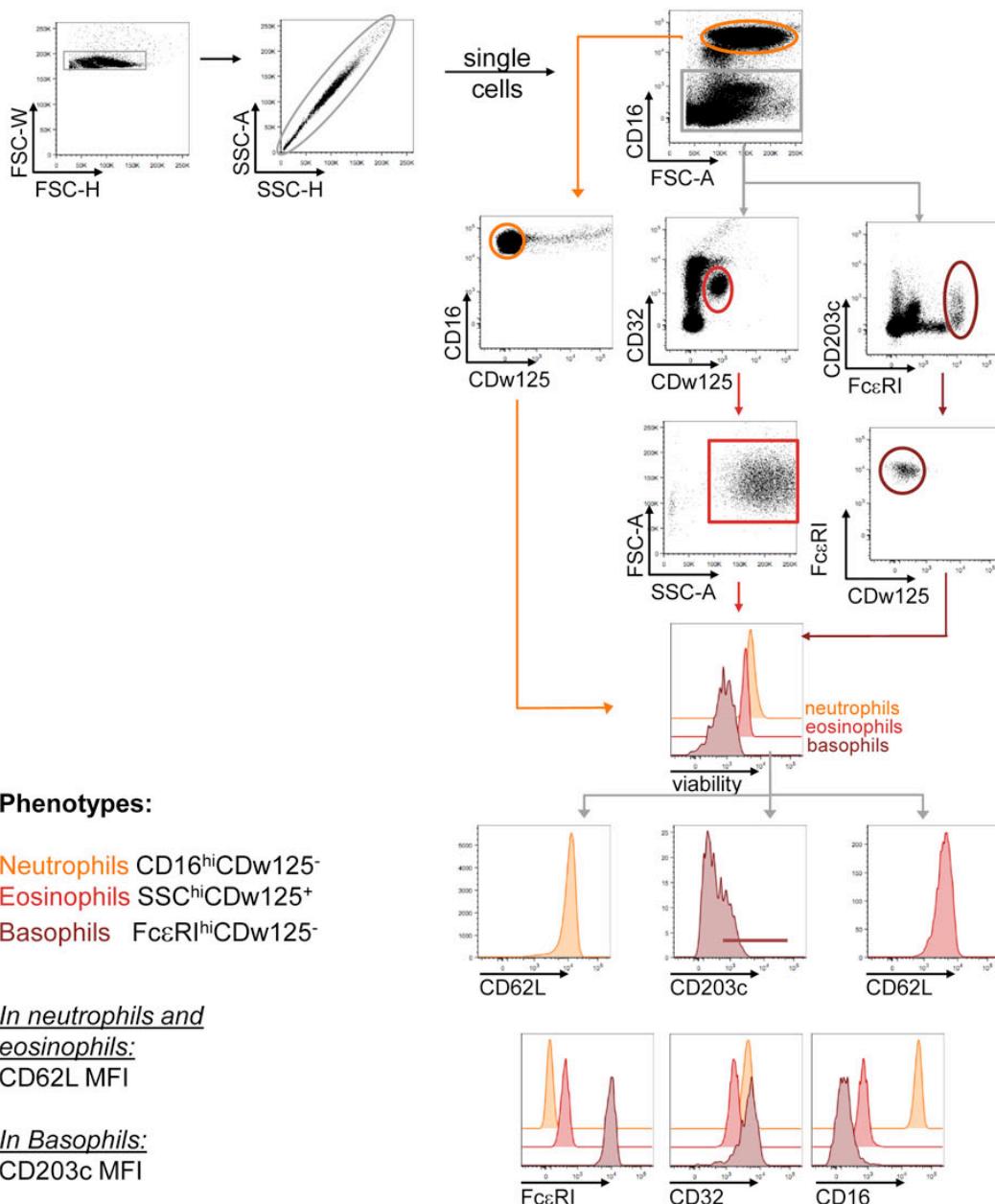
B cell panel



Supplementary Figure 6

Gating strategy for the B cell flow cytometry panel

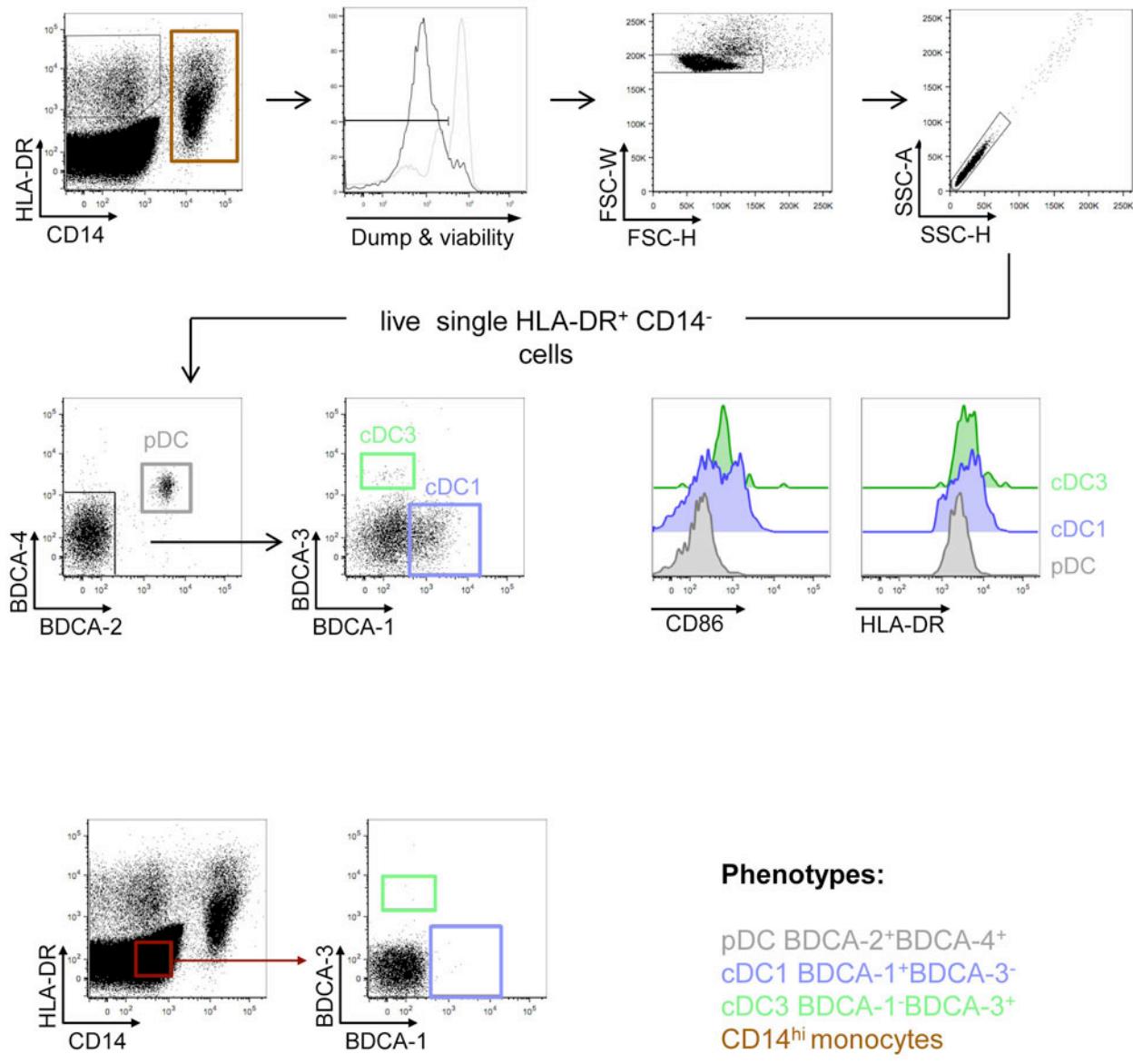
PMN panel



Supplementary Figure 7

Gating strategy for the PMN cell flow cytometry panel

DC panel



Phenotypes:

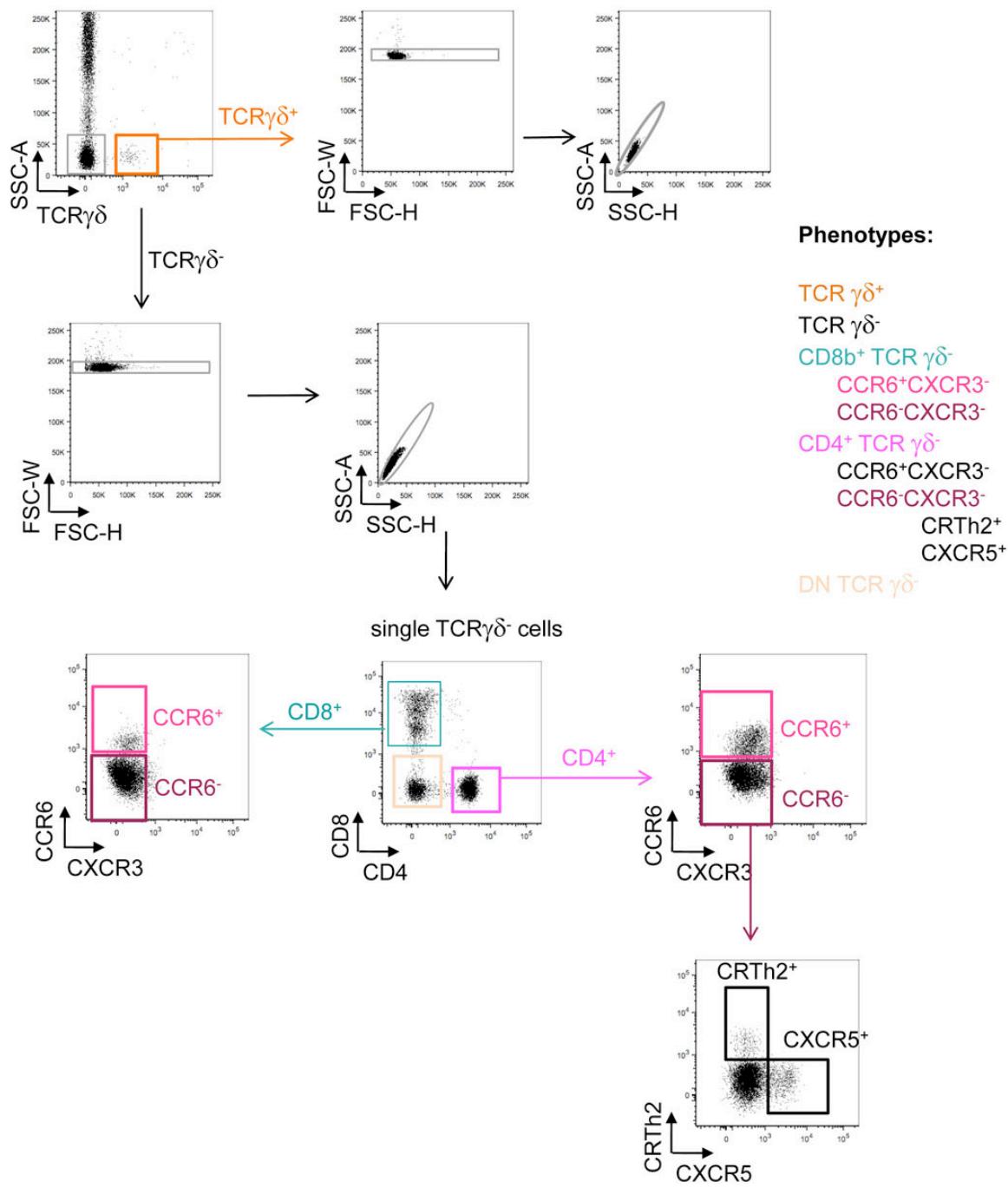
pDC BDCA-2⁺BDCA-4⁺
cDC1 BDCA-1⁺BDCA-3⁻
cDC3 BDCA-1⁺BDCA-3⁺
CD14^{hi} monocytes

In each DC population:
 CD86 MFI
 HLA-DR MFI

Supplementary Figure 8

Gating strategy for the DC flow cytometry panel

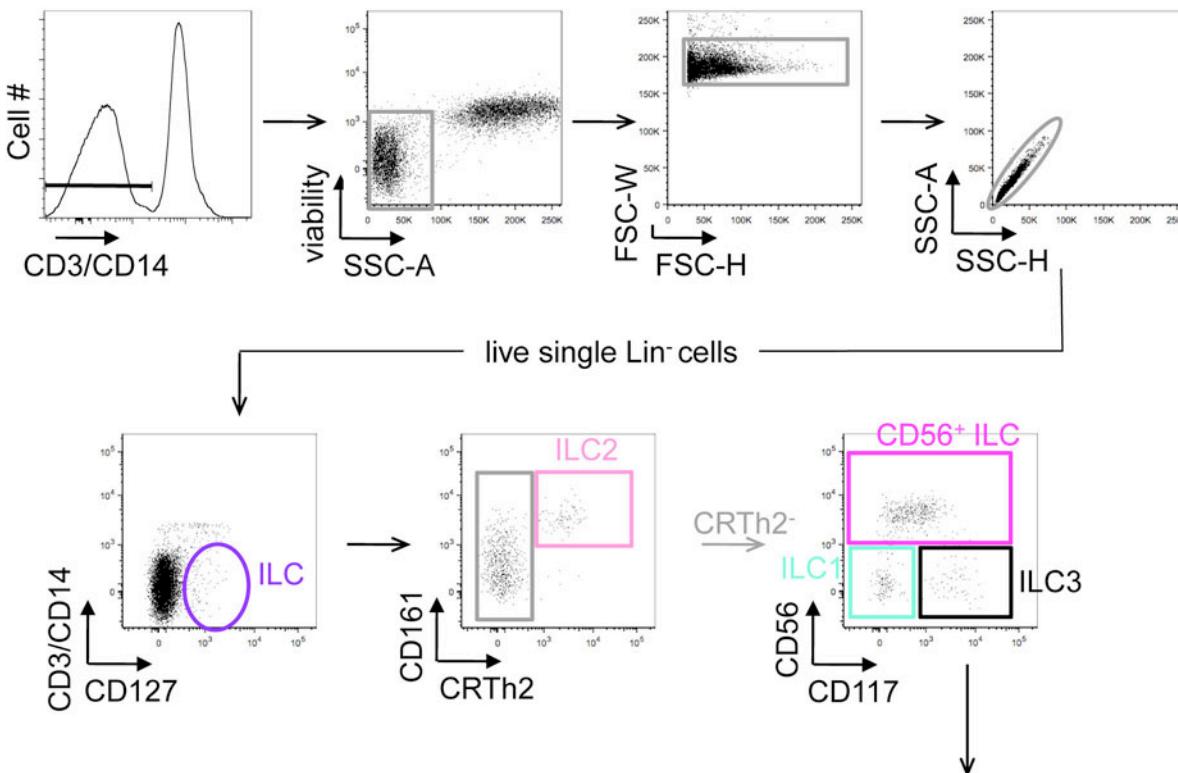
Th panel



Supplementary Figure 9

Gating strategy for the T_H cell flow cytometry panel

ILC panel



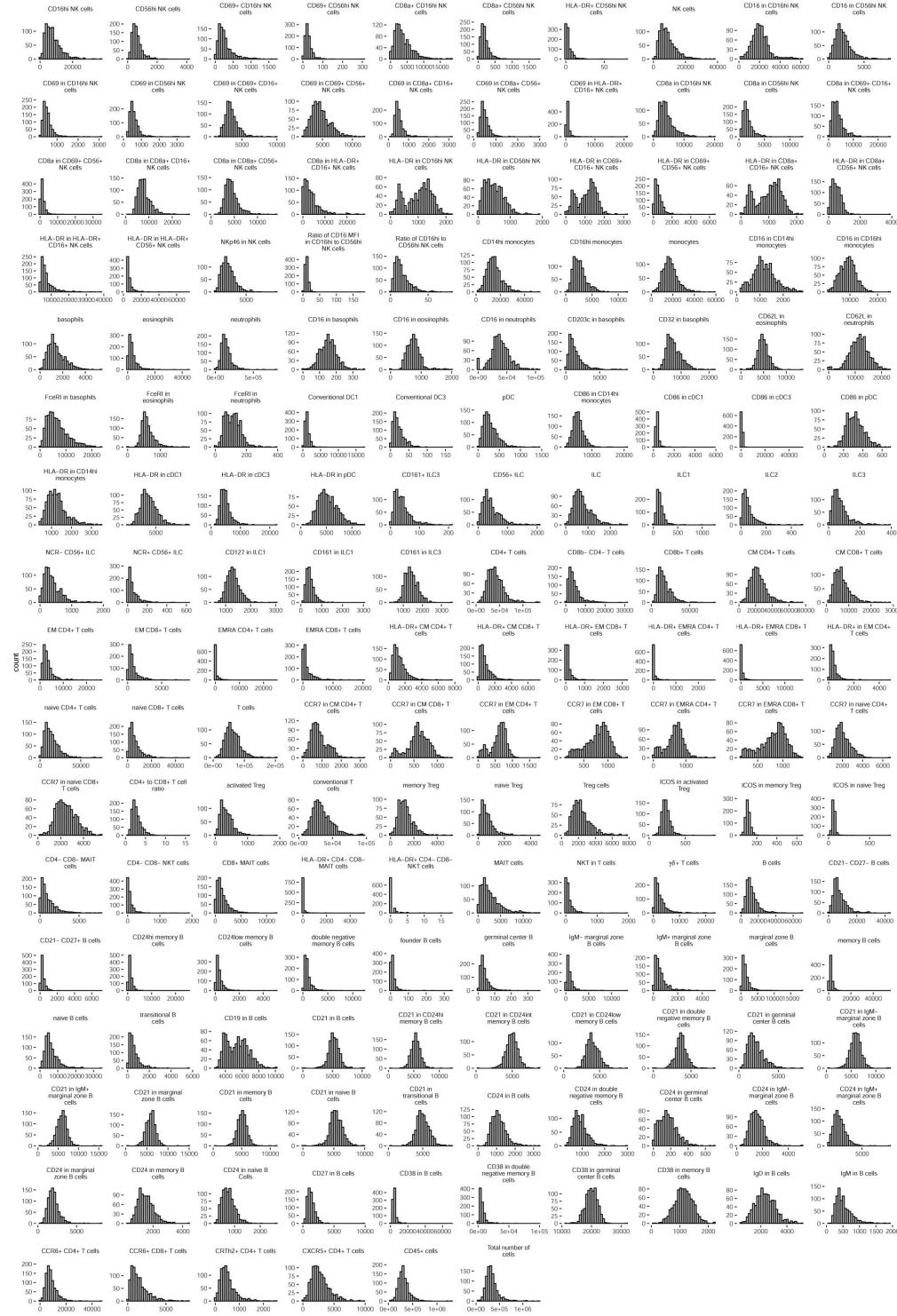
Phenotypes:

ILC (CD127⁺lin⁻)
ILC1 (CRTh2⁻CD56⁻CD117⁻)
ILC2 (CRTh2⁺CD161⁺)
ILC3 (CRTh2⁻CD56⁻CD117⁺)
CD56⁺ ILC (CRTh2⁻CD56⁺)

In ILC3:
CD161 MFI

Supplementary Figure 10

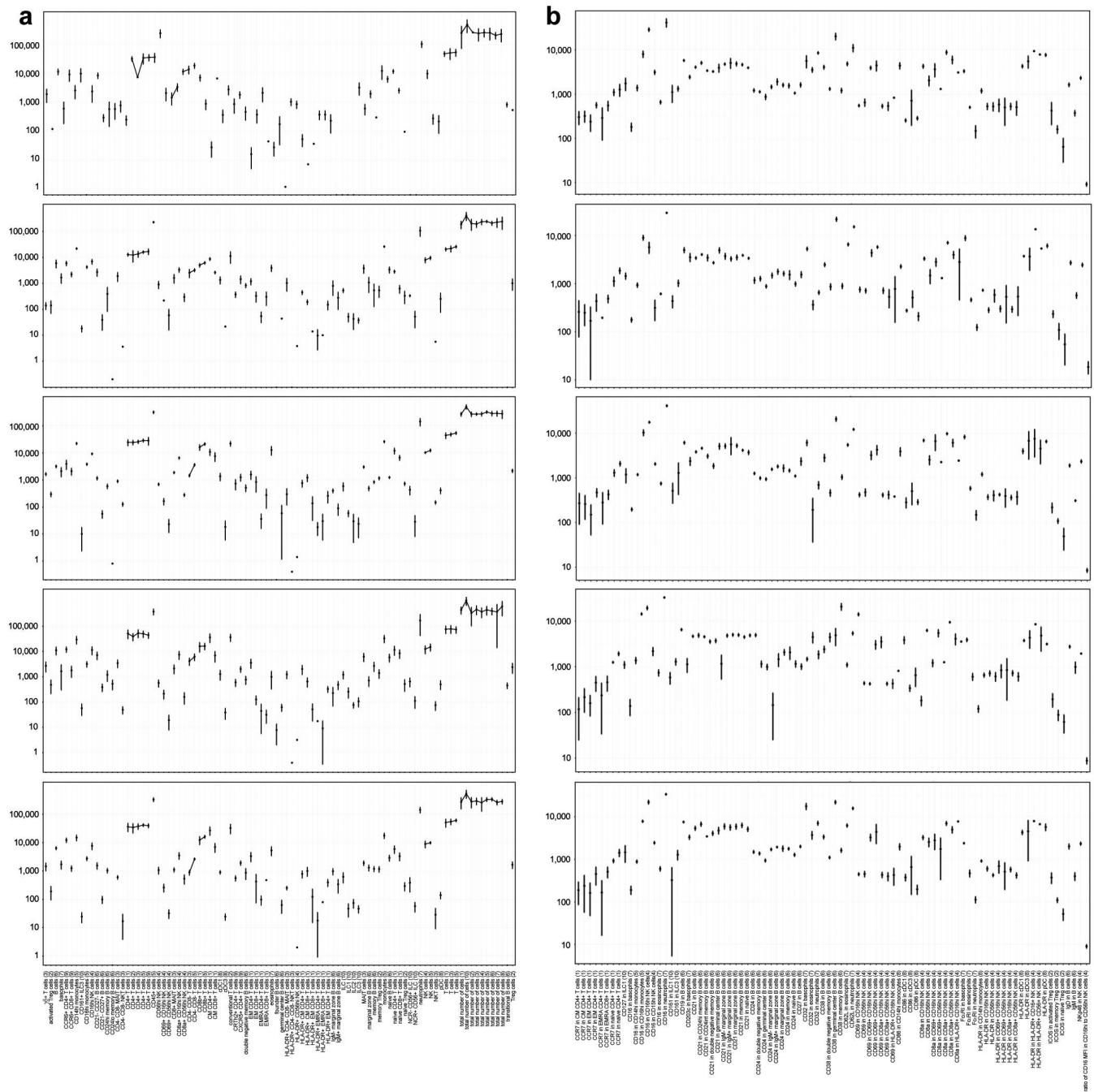
Gating strategy for the ILC flow cytometry panel



Supplementary Figure 11

Raw distributions of the 166 immunophenotypes studied in the Milieu Intérieur cohort

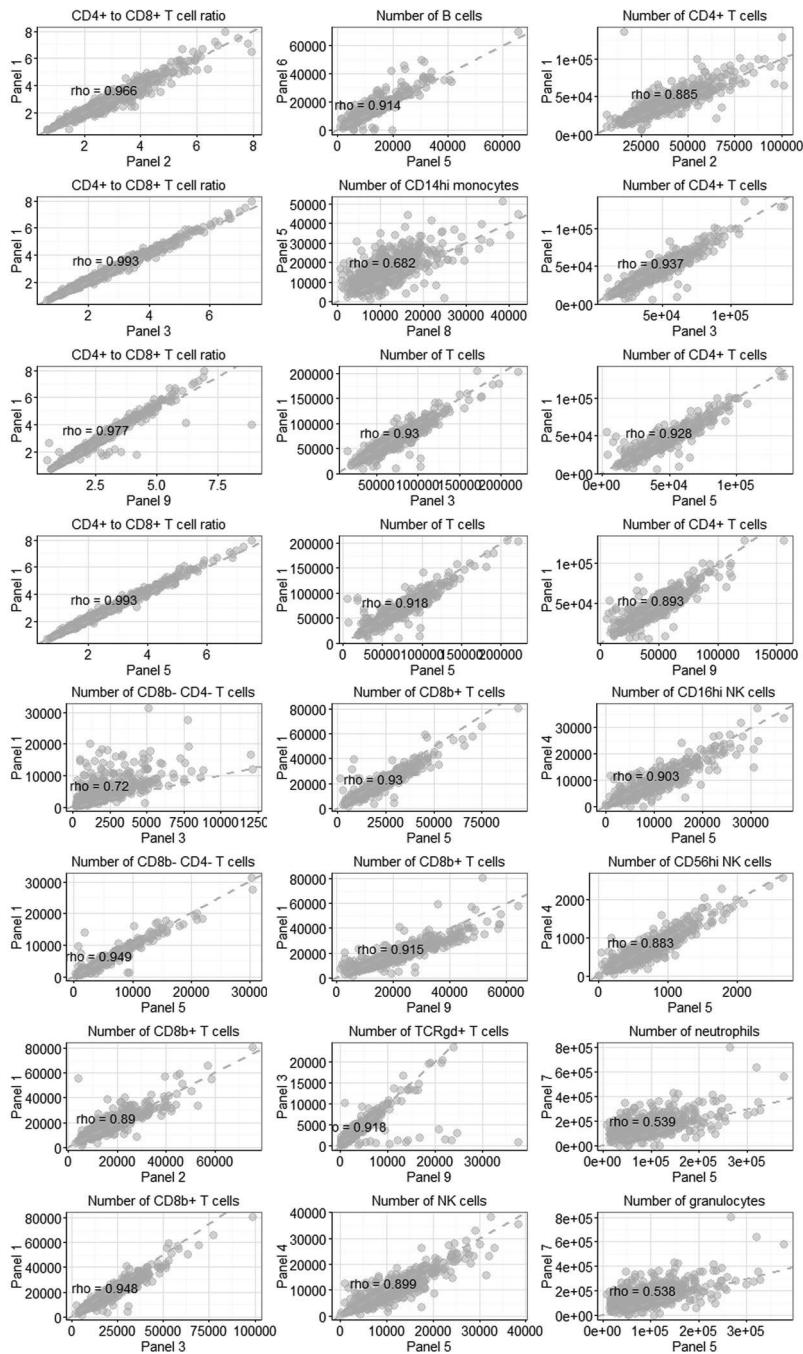
Distributions are shown before outlier removal, imputation, batch correction and transformation to normality. Transformed distributions are shown in **Supplementary Figure 15**.



Supplementary Figure 12

Repeatability of semi-automated flow cytometry measurements

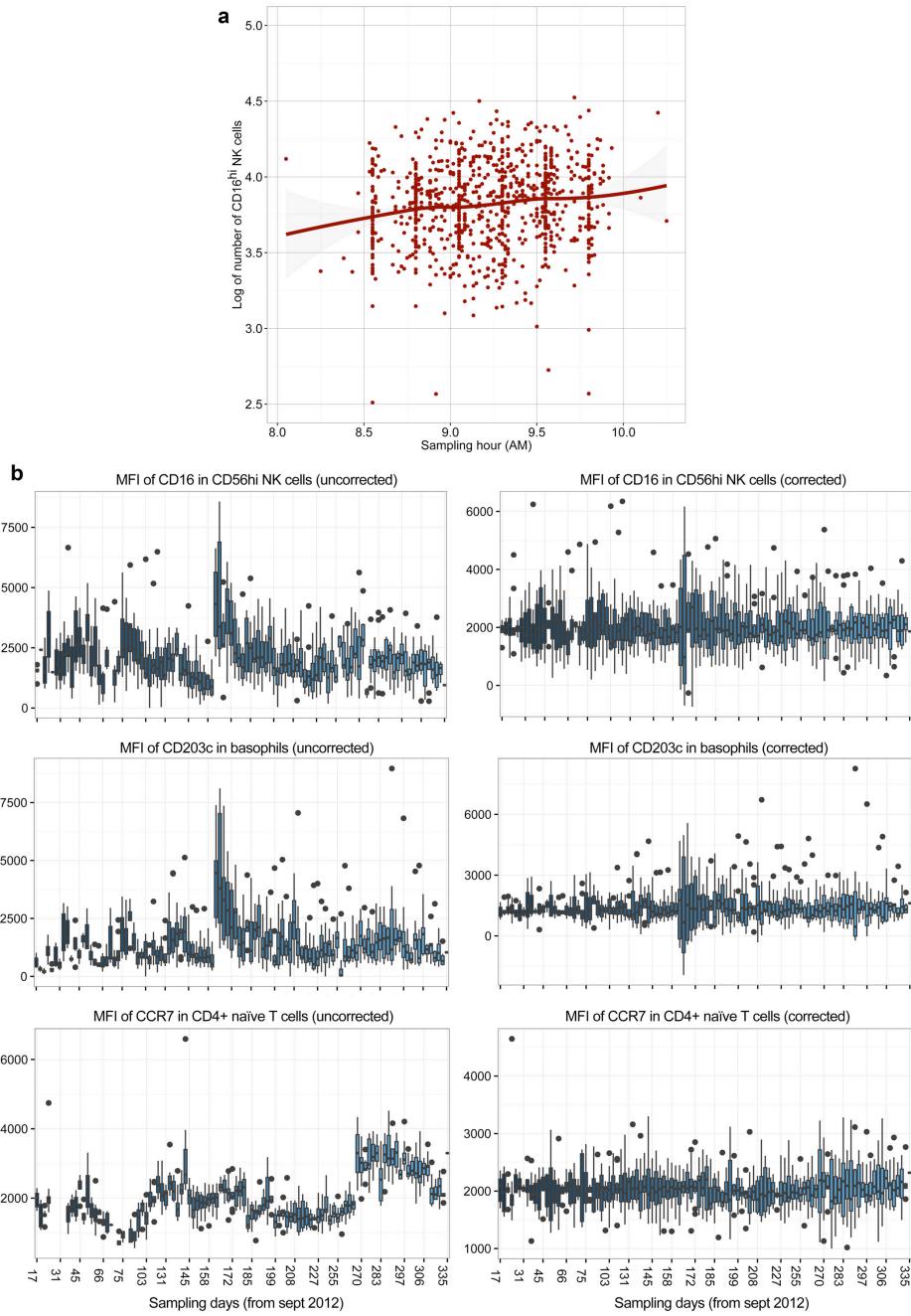
Repeatability of **(a)** absolute cell counts and **(b)** MFI estimated based on measurements performed at five different time points over five months on the same five individuals. Points represent the average of measurements across replicates of the same individual, and bars 3 times their standard deviation.



Supplementary Figure 13

Reproducibility of semi-automated flow cytometry measurements

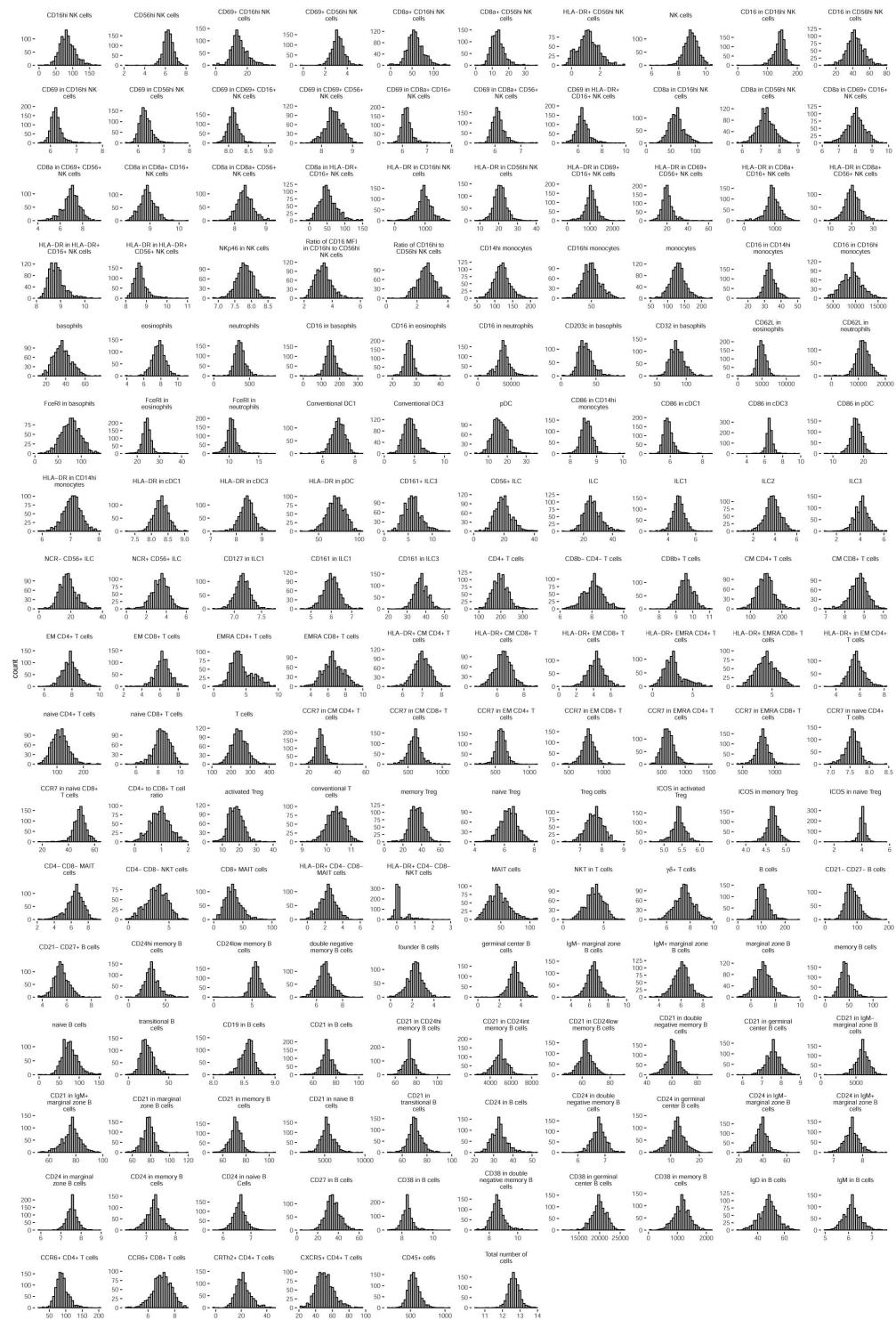
Comparison of cell counts measured in different flow cytometry panels across the 1,000 Milieu Intérieur donors. Spearman's ρ correlation coefficients are shown.



Supplementary Figure 14

Batch effects on measured immunophenotypes

(a) Effect of the sampling hour on the absolute number of CD16^{hi} NK cells. Whole blood samples were collected from the 1,000 Milieu Intérieur healthy subjects every working day from 8 to 11AM. The sampling hour has a significant effect on different immune cell counts, such as the absolute number of CD16^{hi} NK cells, which were adjusted for this batch effect. (b) Effect of the sampling day on immunophenotypes measured in the Milieu Intérieur cohort. Whole blood samples were collected from the 1,000 Milieu Intérieur healthy subjects every working day from September 2012 to august 2013. Variation in the mean fluorescence intensity (MFI) of several protein markers was observed along the sampling period. All MFIs were corrected for this batch effect using the ComBat non-parametric Bayesian framework. Three examples of uncorrected (left column) and corrected (right columns) values are shown.



Supplementary Figure 15

Transformed distributions of the 166 immunophenotypes studied in the Milieu Intérieur cohort

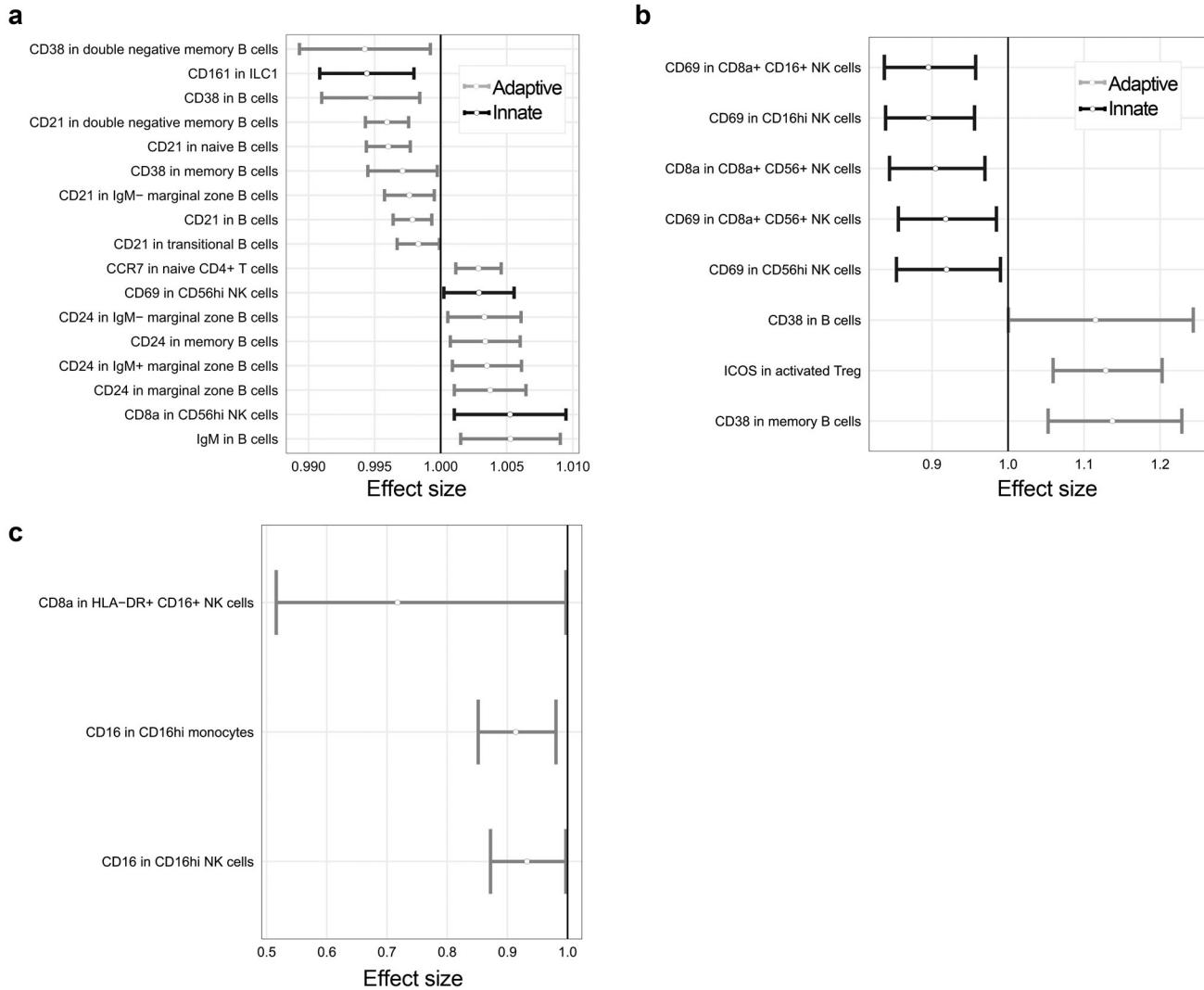
Distributions are shown after outlier removal, imputation, batch correction and transformation to normality. Raw distributions are shown in **Supplementary Figure 11**.



Supplementary Figure 16

Correlations among circulating levels of immune cell populations in the Milieu Intérieur cohort

Sample correlations were estimated between the residuals of the log-transformed immunophenotypes from a regression model that included non-genetic covariates selected with the stability selection algorithm (described below) on the 39 non-genetic covariates, together with the batch variables. Correlations between -0.3 and 0.3 were set to 0, to improve readability.

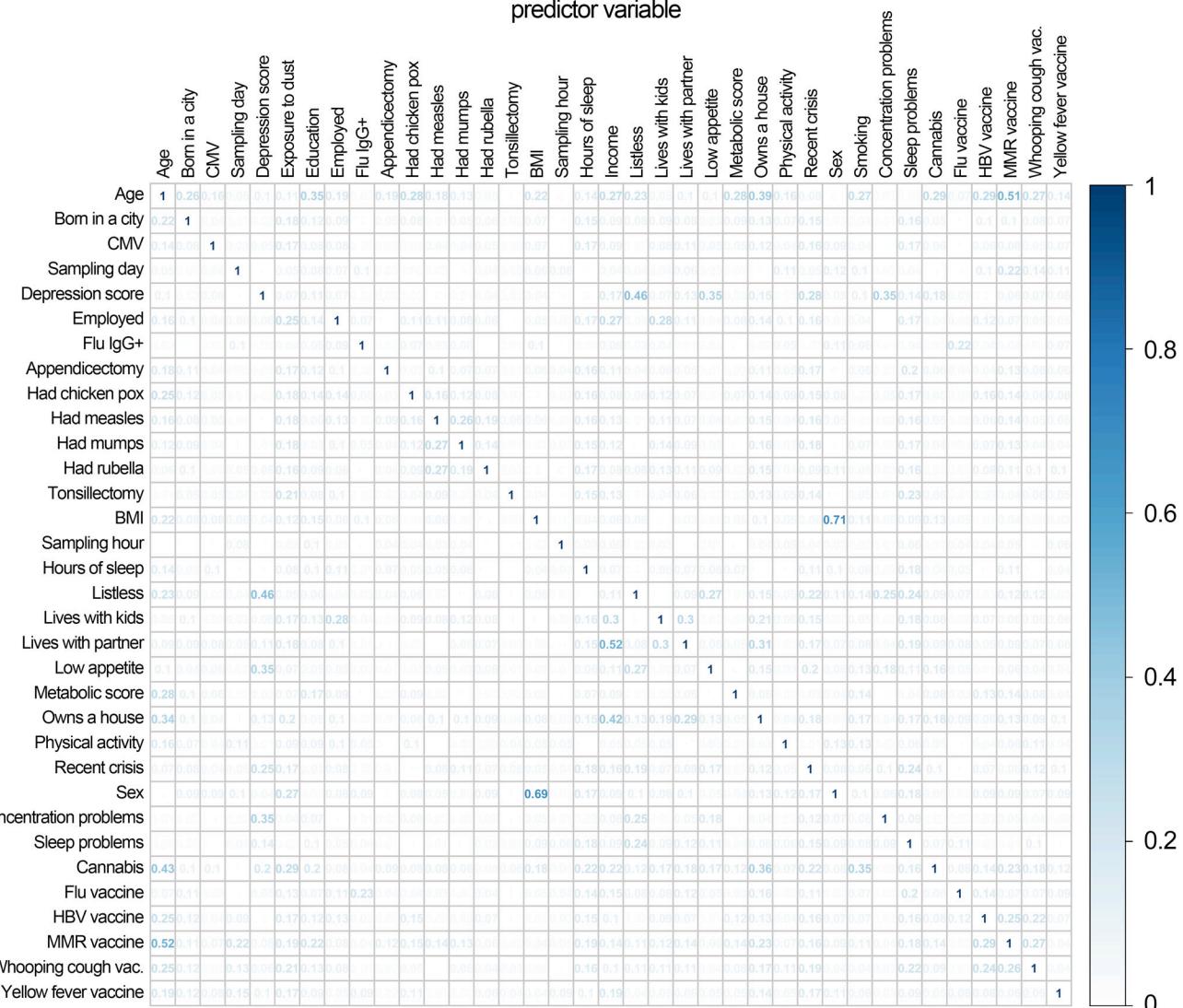


Supplementary Figure 17

Age, sex and CMV infection impact protein levels of immune cell markers in 1,000 healthy individuals

Significant multiplicative effects of (a) increasing age, (b) female sex and (c) CMV seropositivity on the mean fluorescence intensity (MFI) of protein markers of immune cells of 1,000 healthy individuals, controlling for the two other factors, batch effects and genome-wide significant SNPs. The multiplicative effect sizes were estimated in a linear mixed model with a log-transformed response variable, controlling for batch effects and genome-wide significant SNPs, and then transformed to the original data scale. The 99% confidence intervals (99%CIs) were false coverage-adjusted, considering that only 99%CIs of effect sizes with a significant test (adjusted $P < 0.01$) are shown. Adaptive and innate immune cells are represented in grey and black, respectively. These analyses were performed with the *mmi* R package (**Online Methods**).

response variable



Supplementary Figure 18

Estimated dependency matrix among 39 non-genetic variables analyzed in this study

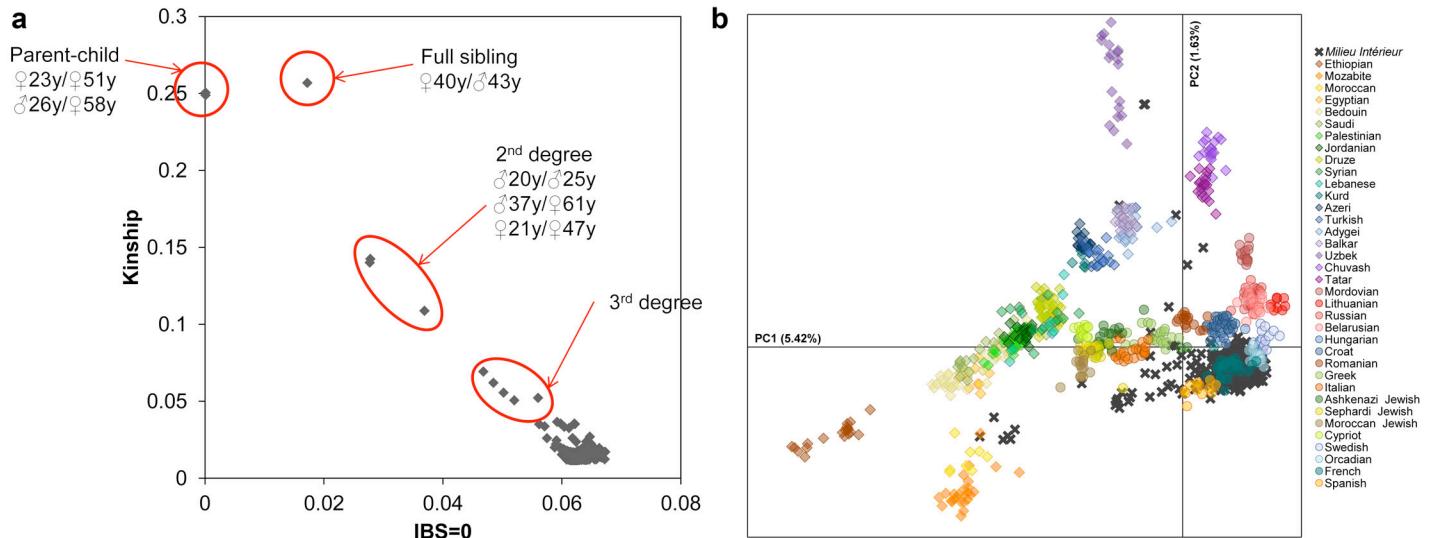
Values are generalized R^2 measures from pairwise fitted generalized linear models (**Online Methods**). Variables listed in rows are response variables and those listed in columns are predictor variables in the models. Categorical variables (*i.e.*, exposure to dust, level of education, income and smoking) were used only as predictors. Details on the non-genetic variables can be found in **Supplementary Table 1**.



Supplementary Figure 19

Smoking strongly impacts protein levels of immune cell markers in 1,000 healthy individuals.

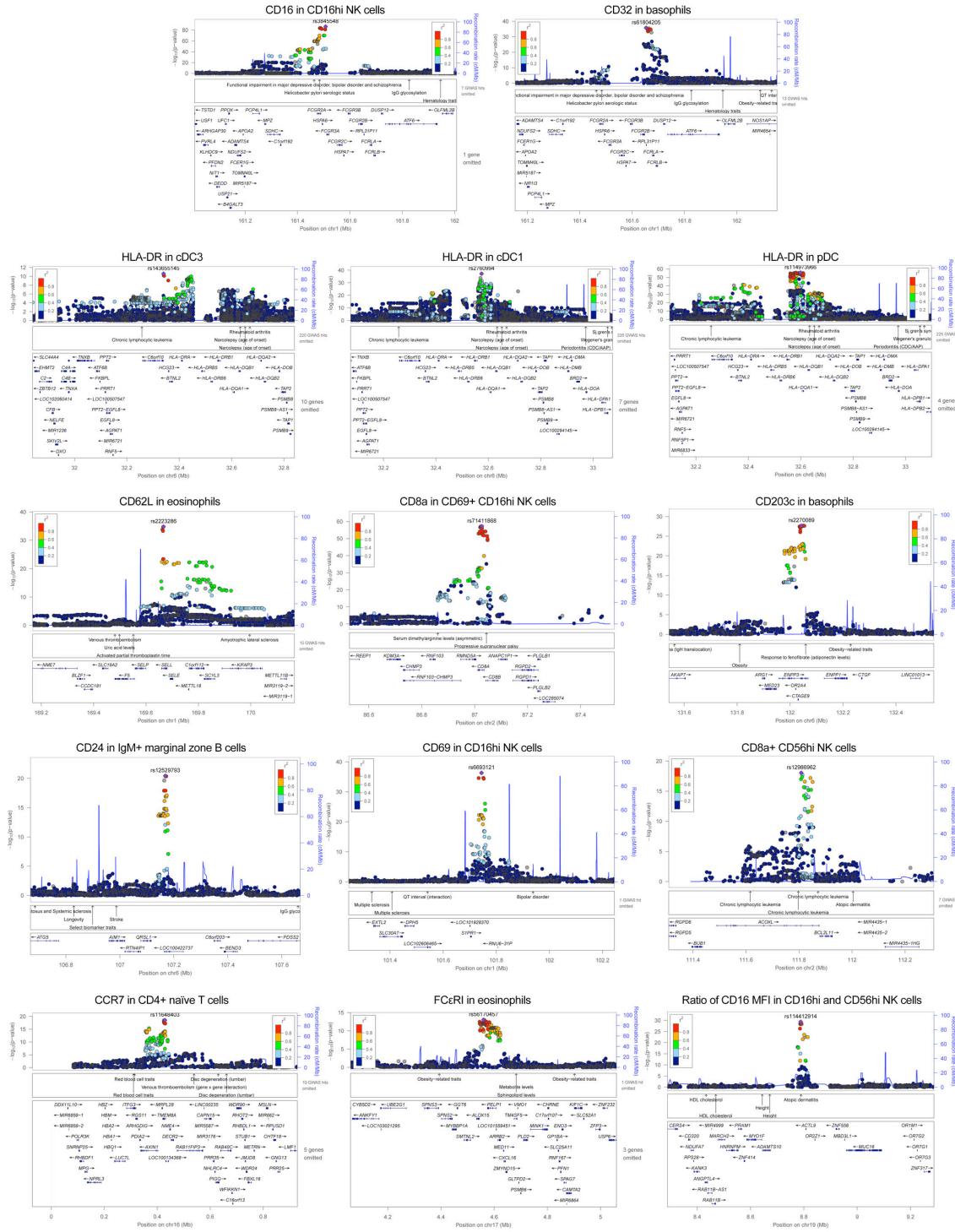
(a) Significant multiplicative effects of 39 non-genetic factors on protein levels of immune cell markers (*i.e.*, MFI) in healthy individuals. Colors represent levels of association (*i.e.*, $-\log_{10}(q\text{-values})$) between the 39 non-genetic factors and protein levels of immune cell markers, at a false discovery rate (FDR) < 1%. Except when their effects were specifically measured, immunophenotypes were regressed for age, sex, CMV status, batch effects and genome-wide significant SNPs (**Online Methods**). **(b)** Significant effect sizes of active smoking on protein levels of immune cell markers in 1,000 healthy individuals. The multiplicative effect sizes were estimated in a linear mixed model with a log-transformed response variable, controlling for age, sex, CMV status, batch effects and genome-wide significant SNPs, and then transformed to the original data scale. The 99% confidence intervals (99%CIs) were false coverage-adjusted, considering that only 99%CIs of effect sizes with a significant test (adjusted $P < 0.01$) are shown. Adaptive and innate immune cells are represented in grey and black, respectively. Effect sizes in past smokers are shown, for comparison purposes. These analyses were performed with the *mmi* R package (**Online Methods**).



Supplementary Figure 20

Genetic relatedness and structure in the Milieu Intérieur cohort

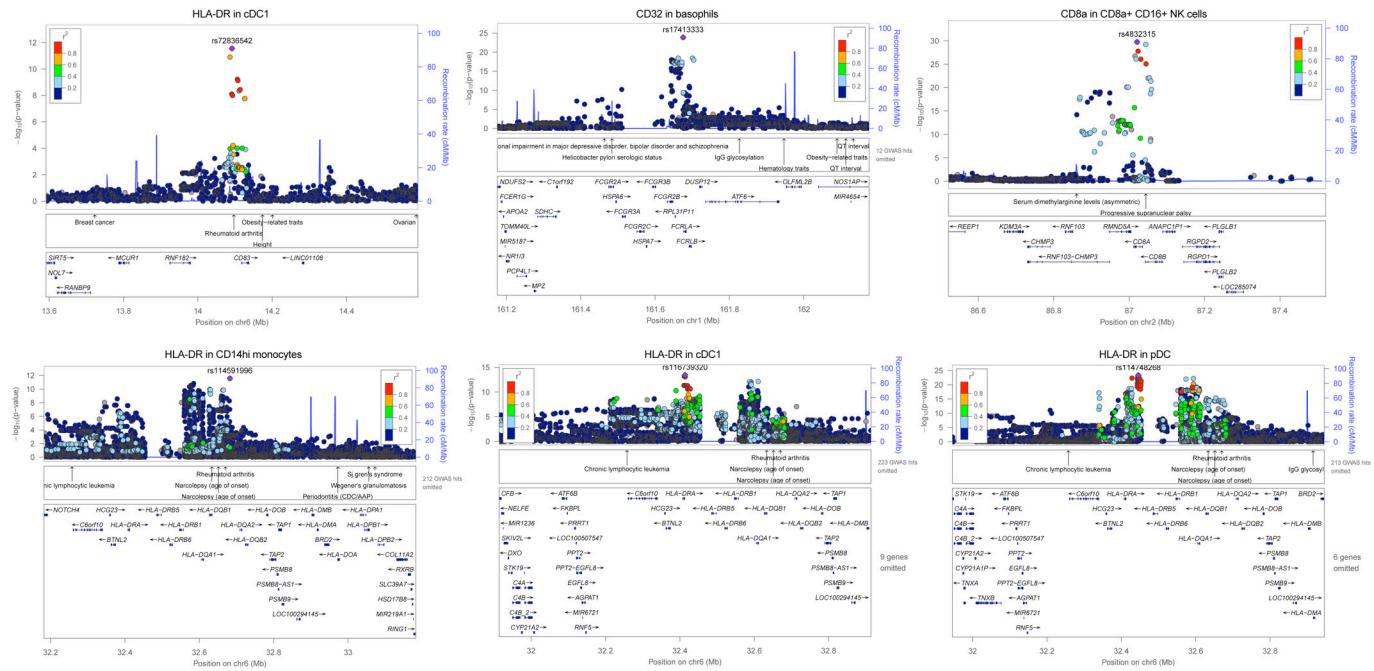
(a) Genetic relatedness in the *Milieu Intérieur* cohort. Pairs of related subjects were detected using an estimate of the kinship coefficient and the proportion of SNPs that are not identical-by-state between all possible pairs of subjects, using KING. (b) Genetic structure of the *Milieu Intérieur* cohort. Genetic structure was estimated with the Principal Component Analysis (PCA) implemented in EIGENSTRAT. For comparison purposes, the analysis was performed on 261,827 independent SNPs and 1,723 individuals, which include the 1,000 *Milieu Intérieur* subjects together with 723 individuals from a selection of 36 populations of North Africa, the Near East, western and northern Europe (**Online Methods**).



Supplementary Figure 21

Local association signals for the 15 genome-wide significant hits associated with immunophenotypes measured in the Milieu Intérieur cohort

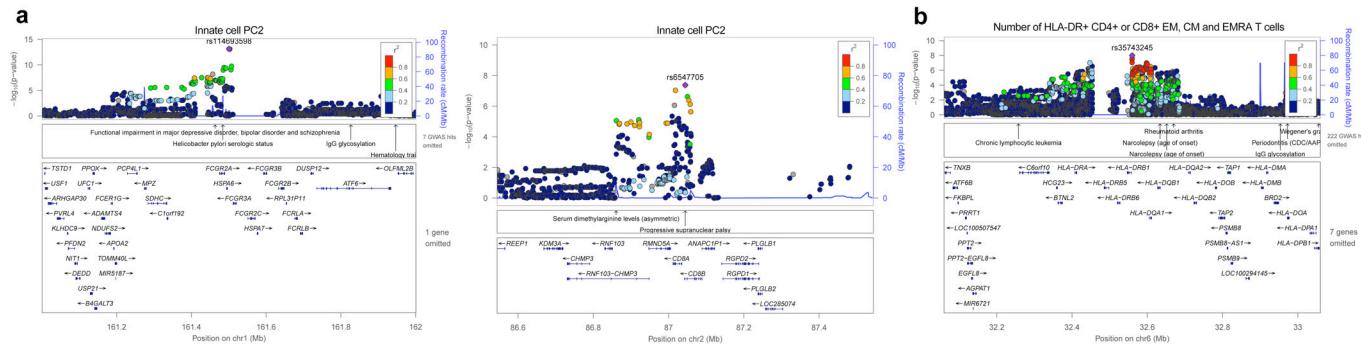
Each point is a SNP, whose color represents its level of linkage disequilibrium (r^2) with the best hit (in purple). Blue lines indicate local recombination rates.



Supplementary Figure 22

Local association signals for the six genome-wide significant hits identified by conditional GWAS of the 15 immunophenotypes showing strong genetic association in the Milieu Intérieur cohort

Each point is a SNP, whose color represents its level of linkage disequilibrium (r^2) with the best hit (in purple). Blue lines indicate local recombination rates. These association signals were identified when conditioning on genotypes of main GWAS signals (**Online Methods**).



Supplementary Figure 23

Local association signals detected by multi-trait GWAS of immunophenotypes measured in the Milieu Intérieur cohort

(a) Local association signals at loci influencing principal components of a PCA of all innate cell immunophenotypes. No suggestive signal ($P < 5 \times 10^{-8}$) was observed for PCs of a PCA of adaptive cell immunophenotypes. **(b)** Local association signal with the absolute numbers of HLA-DR⁺ T_{CM}, T_{EM} and T_{EMRA} cells, either CD4⁺ or CD8⁺. The six corresponding immunophenotypes were analyzed altogether using a multivariate GWAS (**Online Methods**).